BR99599
②

AD-A169 547

RSRE
MEMORANDUM No. 3945

ROYAL SIGNALS & RADAR ESTABLISHMENT

PHYSICS FOR TRAVELLING SALESMEN: SOME NEW
APPROACHES TO COMBINATORIAL OPTIMISATION

Author: D G Bounds

RSRE MEMORANDUM No. 3945

DTIC FILE COPY

PROCUREMENT EXECUTIVE,
MINISTRY OF DEFENCE,
R SRE MALVERN,
WORCS.

DTIC
ELECTED
JUL 1

Memorandum 3945

TITLE: PHYSICS FOR TRAVELLING SALESMEN: SOME NEW APPROACHES
           TO COMBINATORIAL OPTIMISATION

AUTHOR: D G Bounds

DATE: March 1986

SUMMARY

        This Memorandum is a draft version of a review paper which will be
submitted to the Bulletin of the Institute of Mathematics and its
Applications.

        Three new approaches to combinatorial optimisation have been described
recently. They are based on analogies with physical and biological systems.
The first, Kirkpatrick's Optimisation by Simulated Annealing method, has
already proved useful in engineering optimisation problems such as layout and
routing in VLSI chip design. At present, this is probably the most powerful
general optimisation method for use on conventional serial computers. The
second, Brady's evolution-based approach, has yet to receive a thorough
numerical study. However, it seems likely to provide powerful optimisation
methods for parallel SIMD computer architectures of which the most widely
distributed example in the UK is the ICL DAP. The third method, Hopfield and
Tank's Optimisation via networks of amplifiers, is the most revolutionary and
may lead to a new generation of chips with mixed analog/digital functions for
rapid optimisation.

QUALITY
INSPECTED
1

CONTENTS

1. INTRODUCTION

Optimisation problems are ubiquitous: they arise in science and engineering, in industry and commerce, and in the physical world around us. Many, perhaps most, interesting optimisation problems are hard because they correspond to decision problems which are known to be NP complete, i.e. an exact solution requires a number of computational steps that grows faster than any finite power of the size of the problem. A classic example is the Travelling Salesman Problem (TSP) which consists of finding the shortest tour between N cities visiting each once only and ending at the starting point. There are several varients of this problem; the best known (Dutch?) version can be stated more formally: construct the polygon of minimum perimeter through a given set of points in the plane. For N cities, there are $(N-1)!/2$ possible tours. An exhaustive search, which is the only certain way to locate the global minimum, therefore requires evaluating $(N-1)!/2$ alternatives - a number which grows faster than any finite power of N.

However, Sales Managers do not require perfection; only that their staff don't go to Birmingham by way of Beachy Head! Usually a near-optimal solution is satisfactory, if such a solution exists. We shall show later in this paper that there are physical grounds for believing that at least some NP complete problems do have many local optima with values close to the global optimum (see fig. 1 for an example). In these problems there is little to choose between many solutions and any one of them will be a reasonable estimate of the global optimum value. One can take this argument further. In practical optimisation problems the cost function is usually some rough approximation to the real problem and is designed to include, though in an arbitrary way, the main variables on which the solution depends. Since the cost function is arbitrary, it is more important to find good solutions quickl, than to find the global minimum.

$f(X)$

X
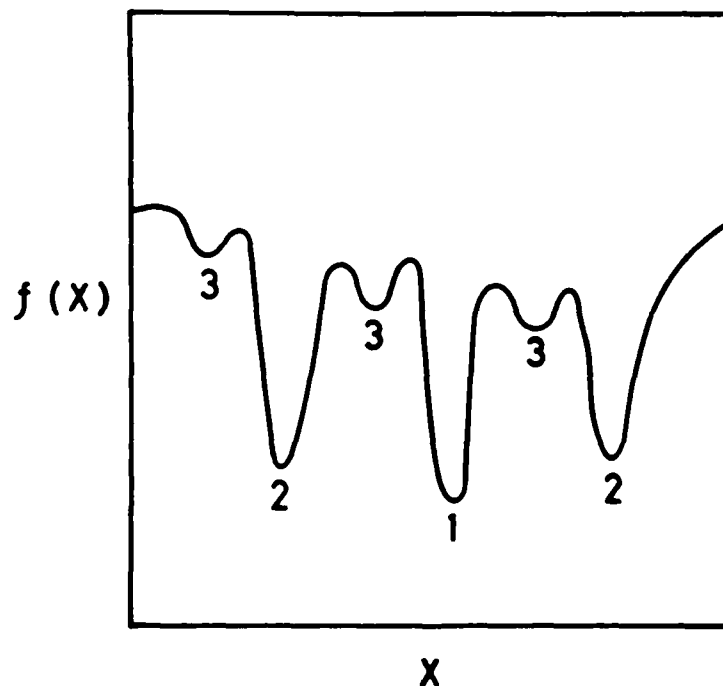
Figure 1: A two-dimensional representation of a cost function $f(X) = f(x_1, x_2, ... x_N)$ which has many local minima with values of f close to the global minimum value. X defines an allowed configuration. In the N-city TSP $f(X)$ is the length of a tour and X is a N-dimensional vector which defines to order that cities occur in a tour; for example the permutation $P = (i_1, i_2, ... i_N)$ where i labels the cities. Point 1 is the global minimum (shortest tour); points 2 are low-cost local minima which are good enough for many practical problems (short tours); points 3 are high-cost local minima from which minimisation methods must be able to escape.

Much of mathematics has been developed in order to describe natural phenomena. I believe that combinatorial optimisation may now benefit from the same source. Three recent pieces of work in apparently unrelated areas of physics seem likely to lead to much better algorithms; and certainly suggest novel ways to think about such problems. The first, optimisation by simulated annealing (OSA), originated in spin-glass physics. The second is based on ideas from biological evolution. The third was stimulated by neurophysiology and models of neural networks. One aim in writing this article is to bring this novel work to the attention of mathematicians in the hope that it stimulates further rigorous developments. This paper is really about combinatorial optimisation and not TSP's per se: however the TSP is a well characterised problem and is therefore convenient for investigating the new methods.

2. OPTIMISATION BY SIMULATED ANNEALING: THE SPIN-GLASS PROBLEM

Historically, the origins of OSA lie in spin-glass physics. Although the method has now been developed to the point where it is no longer necessary to understand something of the spin-glass problem first, it is still useful to do so because spin-glass techniques may yield useful information about combinatorial optimisation problems in addition to the OSA algorithm.

A spin-glass is simple in concept. Consider a lattice of magnetic ions. At high temperatures the thermal energy is sufficient to overcome the interactions between magnetic moments which tend to make the spins align. The spins are therefore able to rotate; the lattice has no net magnetic moment, and the system is paramagnetic. Below a critical temperature, the Curie temperature, the thermal energy is not large enough to overcome the magnetic interactions: the system then adopts an ordered ferromagnetic phase with a net magnetic moment. However, if the magnetic interactions are weakened by replacing most of the magnetic atoms by atoms of a non-magnetic metal, a different behaviour occurs. At low temperatures the thermal energy is not sufficient to prevent spins freezing into particular orientations, but the magnetic interactions are too weak to force the long-range order of a ferromagnet. This frozen disordered state is believed to be the structure of the spin glass phase.

Despite this simple picture, the theoretical treatment of spin-glasses is an active field of research where most of the major issues have only just become tractable. Progress has been reminiscent of peeling onions: as one layer of difficulty is removed new and deeper questions appear. It has taken several years to obtain a satisfactory mean field theory of one of the simplest models, the Sherrington-Kirkpatrick infinite range model, and some unusual new physics has emerged in the process. For optimisation problems the relevant features are these.

The Hamiltonian of a system of N spins is given (in the Sherrington-Kirkpatrick model) by

$$H = -\sum_{i=1}^{N} \sum_{j=i+1}^{N} J_{ij} \, s_i \, s_j \qquad (1)$$

where the spin variables s_i take values ± 1 and the interaction J_{ij}'s are Gaussian random variables. The energy surface defined by eqn (1) is similar to that shown in figure 1. There are many minima of similar energy separated by barriers whose height increases with N. For large N, a system prepared with some set of J_{ij}'s, that is in the region of one energy minimum (an energy "valley"), requires a very long relaxation time to move to a different valley. In the thermodynamic limit when $N \rightarrow \infty$, the barriers become infinite; the energy valleys are termed "pure phases"; and the system is non-ergodic (the ensemble average of any observable is not the same as the time average). There are two key features which lead to multiple minima in spin glasses: disorder and frustration. The disorder arises because there is no long range order of spins on the lattice sites. Frustration [1] is easily understood from figure 2 which shows a section of a two-dimensional square lattice with spins at each vertex, nearest neighbour interactions only, and $|J_{ij}| = 1$. The signs on the bonds are the sign of J_{ij}.
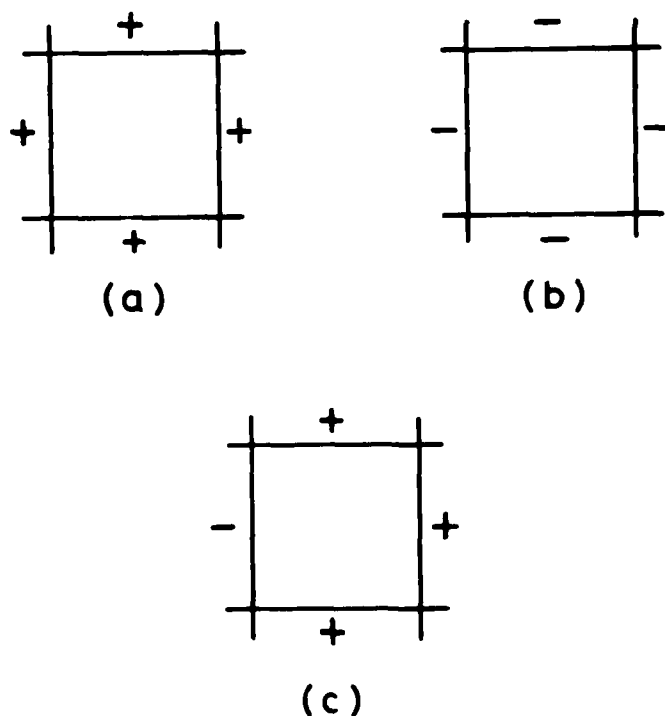

(a)   (b)

(c)

Figure 2:   Section of a 2-dimensional lattice with spins s_i (not shown) at each vertex. The sign on the bonds show the signs of J_{ij} . (a) is a ferromagnetic combination of J_{ij} ; (b), an antiferromagnetic combination; (c), a frustrated (spin glass) combination.

From eqn (1), the lowest energy configuration for fig. 1(a) is when all s_i are equal, which is a ferromagnetic state. For fig. 1(b) the lowest energy is obtained when $s_i = -s_j$; an anti-ferromagnet. However, for fig. 1(c) there is no combination of spins which makes all energy terms negative; at least one bond must be "frustrated". A two-valued frustration function can be defined [1] on closed contours (c) as

$$\phi = \prod_{(c)} J_{ij} \qquad (2)$$

and the system is frustrated if $\phi = -1$ for any closed contour. In general frustration occurs in any system where there are local interactions which favour incompatible types of ordering.

This picture would probably not have emerged without extensive numerical simulations which have guided the development of analytic theories. The principal tool for numerical work is the Monte Carlo method due to Metropolis et al. [2] for estimating ensemble averages such as the expectation value of eqn(1) at some temperature T. A central question in these systems is whether or not some freezing transition to a spin glass phase really occurs as the temperature is lowered. It was natural, therefore, to carry out Monte Carlo calculations with Metropolis sampling at a range of temperatures. A time-varying temperature T(t) can be viewed as an "annealing schedule". Kirkpatrick and co-workers [3] have proposed the same method as a general optimisation technique to find the minimum values of arbitrary cost functions, and the method is now known as Optimisation by Simulated Annealing (OSA). It has been used successfully for a number of combinatorial optimisation problems arising in computer chip design, computer vision, and the TSP. The method was proposed independently by Černý who used it for the TSP [4]. The connection* between the Metropolis method and minimisation was first noted by Pincus [5]*.

OSA operates thus. An optimisation problem involves a cost function $f(X)=f(x_1,x_2,...,x_N)$ to be minimised and a set of candidate solutions generated by trial moves. Gradient methods such as steepest descent accept only those moves which reduce the cost function. Such algorithms have no way to escape from local minima, which is not the behaviour required. To avoid this one can choose new configurations randomly and take the lowest value found after a large number of choices. Randomising algorithms accept moves that results in higher values of the cost function, but these are accepted indiscriminately. Since the probability of finding a near-optimal solution is proportional to the number of near-optimal solutions divided by the total number of possible solutions, randomising algorithms perform less well as the

---

* footnote: I am grateful to B Ripley for bringing this work to my attention.

dimensionality of the search space increases. In contrast, OSA also allows some moves which increase the cost function, but in a controlled manner. If a trial move decreases the cost function, it is accepted. If a trial move increases the cost function, it is accepted with probability $\exp[-\Delta f(X)/T]$ where $\Delta f(X)$ is the increase in the cost function and T is a control parameter (Boltzmann's constant times the temperature for a spin glass - hence the term "temperature" in the OSA literature regardless of what cost function is involved).

A system evolving under these rules eventually reaches thermal equilibrium at any given temperature, and the relative probabilities of two global states α and β is then given by the Boltzmann distribution

$$(P_\alpha/P_\beta) = \exp[-\{f(X_\alpha)-f(X_\beta)\}] \qquad (3)$$

where P_α is the probability of being in state α of cost $f(X_\alpha)$. At high temperatures the probability of accepting uphill moves is greater and equilibrium is reached more quickly. At low temperatures equilibration takes longer but the system is more heavily weighted towards low cost states. The strategy therefore is to do a coarse search of the space at high temperature and then reduce T to focus on the low cost states. Note that OSA is unlike gradient descent algorithms in that it does not find a minimum and then stay there. The algorithm merely spends longer nearer the minimum as T is lowered. To stay in the global minimum would require T=0 in which case it would take an infinite time to reach equilibrium. The T$\rightarrow$0 limit cannot be reached, therefore. However, it is possible to anneal to a low temperature and then use gradient descent to locate the nearest minimum more precisely.

Apart from designing a suitable cost function, the most difficult part of OSA is to choose an annealing schedule which ensures that the system has a Boltzmann distribution of states at low temperatures. It is obvious that this is required if one is modelling a classical liquid or solid; and in learning algorithms such as the Boltzmann Machine [6] which depend upon the mathematical properties of the Boltzmann distribution. However, it is no less necessary for other cost functions. In order to achieve a Boltzmann distribution, it is necessary to keep the system close to equilibrium throughout the annealing process. Therefore it is necessary to reach, or at least come close to, equilibrium at each temperature, and to change the temperature slowly through regions where large decreases in the cost function are observed. The rate of cooling can be quite critical, as the physical analogy on which OSA is based suggests. As metallurgists know well, it is necessary to cool a melt slowly near the freezing temperature in order to grow near-perfect crystals (which are global minimum states). Too rapid cooling results in amorphous structures (metastable states), while lowering T too slowly may result in supercooling (the system approaches the ground state very slowly). The first approach adopted by statistical physicists, notably Kirkpatrick, was to apply methods from condensed matter physics. There it is useful to monitor the specific heat, C(T):

$$C(T) = var(E(T))/k_B T^2 \qquad (4),$$

where k_B is Boltzmann's constant. An increase in $C(T)$ signals the onset of a
phase transition whence a slower cooling rate is necessary. This approach was
used in the original work [3]. More recent practice is described in [7].
However, statisticians have attempted a more rigorous approach to determining
the optimum annealing rate. Geman and Geman [8] provided the first
convergence proof for the algorithm in a paper on image processing.
Unfortunately, the annealing schedule which guarantees convergence is too slow
for practical applications. Gidas has also investigated the conditions under
which the annealing algorithm converges and has proposed a rigorous procedure
for choosing the optimum schedule [9]. Unfortunately for the majority of
people using OSA, who are not statisticians, the results are contained in
sixty pages of mathematics; a fact which suggests that the Kirkpatrick
procedure is likely to be widely used for some time.

In summary, the great advantages of OSA are simplicity and generality.
Though it may need ingenuity to choose an appropriate cost function and move
set, the method is simple to implement and it has been used successfully on a
variety of problems. The main disadvantage is that choosing an annealing
schedule for practical purposes is still something of a black art.

Several studies of the travelling salesman problem have been carried
out using OSA [3,4,10,11]. In addition to choosing an annealing schedule, it
is necessary to define the set of moves allowed. Most studies of the TSP make
use of a move set suggested by Lin [12]. A tour is specified by the
permutation $P=(i_1, i_2, \ldots, i_N)$ with an associated length $L = d_{i_1 i_2} + d_{i_2 i_3} + \ldots + d_{i_N i_1}$
where d_{ij} is the distance between cities i and j. Lin defines a tour to be
"λ-OPT" if no shorter tours can be obtained by replacing λ steps of the tour
(bonds) with any other set of λ bonds. This provides a natural set of
increasingly restrictive constraints to satisfy since:

(1) any tour is 1-OPT;

(2) a tour is optimal if and only if it is N-OPT;

(3) a λ-OPT tour is also λ'-OPT for $\lambda' < \lambda$ [12].

This suggests that one simple move is the interchange of two bonds. Two steps
$d_{i,i+1}$ and $d_{j,j+1}$ on the tour are replaced by d_{ij} and $d_{i+1,j+1}$ as
illustrated in figure 3. A 2-OPT tour is optimal w.r.t all such interchanges.
Most of the annealing studies have used two-bond moves, and it is found that
for N>50 OSA compares favourably with exhaustive two-bond and three-bond
searches, and is progressively better at finding short tours as N increases.
This is shown in figure 4 which is reproduced from Kirkpatrick and Toulouse
[10]. For N>250 it has been concluded that OSA is the best procedure known
[11]; though this statement was made before the methods outlined in sections 3
and 4 were presented! Whether this remains true is a matter for further
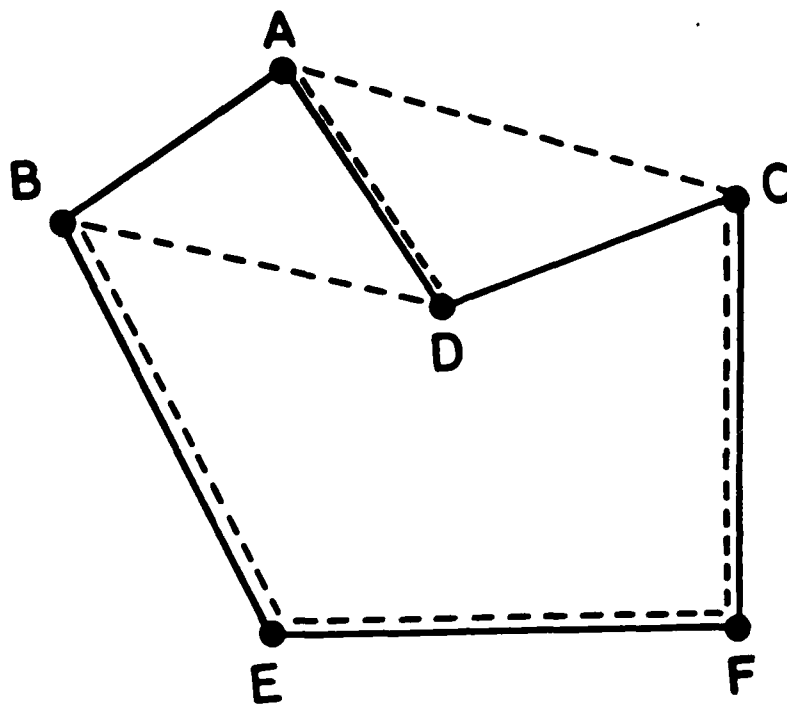research.

Figure 3: A two-bond move which converts the dashed tour ADBEFC into the solid tour ABEFCD.
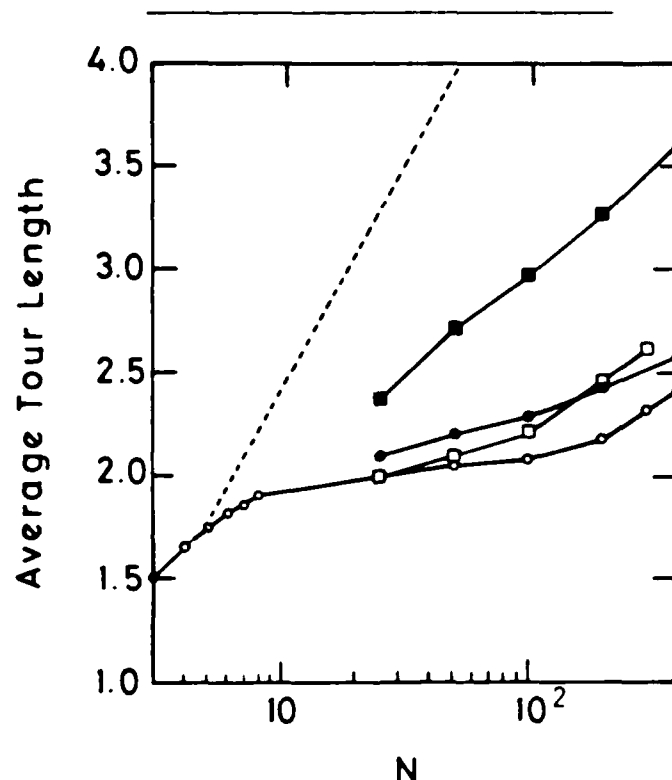

Figure 4: TSP results from Kirkpatrick and Toulouse [10]. This figure shows the optimal tour length for random distance TSP's of up to N = 400 cities, as obtained from various algorithms. The dashed line is the upper bound provided by the greedy algorithm ("Choose a city at random. The next city on the tour is the closest one"). The solid points are from iterative improvement using exhaustive search for 2-bond rearrangements (squares) and 3-bond rearrangements (dots). The open points for N > 24 are OSA results using 2-bond moves (squares) and 3-bond moves (circles). The open data for N < 12 are exact results. Reproduced with permission of the publisher.

However, in addition to providing a useful simulation tool, it seems likely that spin glass physics will reveal much deeper insights into combinatorial optimisation problems. The most exciting results are yet to come. Just as in the spin glass problem numerical studies led to new analytic results, so a deeper understanding of the TSP is beginning to emerge via spin glass techniques. First, we note that the problem of finding the ground state of a 3-dimensional spin glass and the TSP are now known to be of the same computational complexity, since Bachas [13], building on earlier work of Barahona [14], has shown that the former problem is NP complete. This provides some justification for the application of spin glass methods to other NP complete problems. Second, several workers have formulated the TSP in statistical mechanical terms. Orland [15] has presented a mean field version of the TSP, and has conjectured that NP completeness is associated with a concept called replica symmetry breaking which is well known in spin glass physics where it was first introduced by Edwards and Anderson [16]. A connection between replicas and P class optimisation problems has also been suggested by Mezard and Parisi [17], and their results have recently been extended to the NP case [18]. When the TSP is transformed to a statistical physics problem, quantities such as the average length of tours become thermodynamic quantities; and one is interested in their behaviour as a function of temperature. Vannimenus and Mezard [19] have shown the existence of two different temperature regimes which exhibit quite different behaviour in the way such properties vary as a function of N. They have also obtained the exact asymptotic behaviour of the tour of optimal length; though only for a TSP in infinite dimension! Kirkpatrick and Toulouse [10] have introduced a simplified version of the TSP in which the symmetric distances ($d_{ji} = d_{ij}$) are random variables in the interval $(0,1)$. The reason for studying this model is that it may prove analytically soluble, as did the Sherrington-Kirkpatrick model for spin glasses. I can do no more here than point to the rapid progress in this area; for a deeper discussion of the relationship between spin glass phenomena and TSP's the reader is urged to consult the pioneering paper of Kirkpatrick and Toulouse [10]. Among the tantalising possibilities is that a new, more refined characterisation of computational complexity may emerge. Some work has been reported in this direction [20].

3. OPTIMISATION STRATEGIES FROM BIOLOGY

Brady has recently suggested a different approach to the TSP motivated by the idea that biological evolution overcomes the local "fitness maxima" which correspond to stagnant species [21]. Since evolution has been going on for a long time, one would expect to find efficient evolutionary strategies in modern flora and fauna. A study of those mechanisms might therefore be useful for other optimisation problems.

Consider an arbitrary tour of N cities. In the biological terms adopted by Brady, a random alteration to the tour is identified as a mutation; and simulated annealing may be viewed as a method for accepting unfavourable mutations with non-zero probability. Brady further restricts his study to local mutations which consist of swapping the order of two cities in the tour. Such mutations belong to a subset of the set of two-bond moves, originally suggested by Lin [12], which were utilised in the annealing studies. The baseline on which all other methods improve (Brady's method A, results shown in figure 5) is to choose trial two-bond moves at random but accept only those which shorten the tour. This single random quench is merely an inefficient way
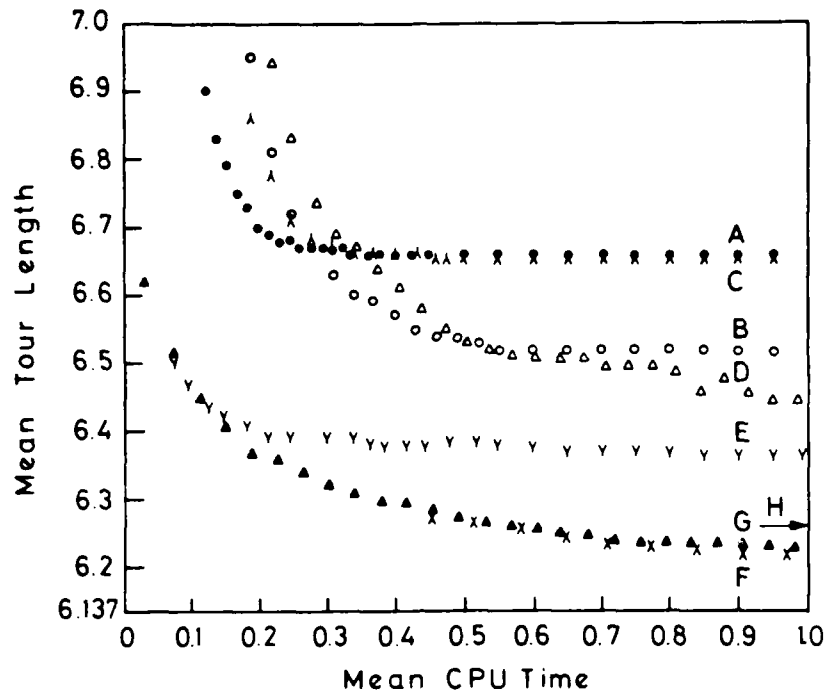
Figure 5:   Brady's results for the TSP [21]. A comparison of eight different optimisation strategies. In each case, data were averaged over 100 separate optimisations. A, quenching by random attempted mutations. B, the best of two independent quenches. C, competition between two quenches. D, mating between two quenches. E, mating between two quenches obtained by systematic search. F, mating between 12 quenches obtained by systematic search. G, best of n independent quenches obtained by systematic search, n = 1, 2, 3,..., 26. H, simulated annealing; 20,000 random attempted mutations per temperature (statistics from 20 optimisations only, with mean time 5.6 s). Reproduced by permission from [21]. Copyright (C) 1985 Macmillan Journals Ltd.

---

to find the nearest local minimum. Brady suggests that a lineage developing in a similar manner would be at an evolutionary disadvantage since it is destined to stagnate and be unable to adapt to changing circumstances. Prospects of survival are improved by diversification into several independent species, at the cost that evolution may be slower since resources must be shared. For optimisation, this suggests dividing the available computer time into two or more independent starting tours and quenching each separately (method B). As expected, this produces shorter tours on average than a single random quench. The effect of competition between species may be modelled by taking two independent tours, carrying out random quenches, but every τ attempted mutations the longer tour is discarded, a second copy of the shorter tour is made and the quenches are continued (method C). Since this forces the two lines back to a common point, it is not surprising that on average this method performs no better than a single random quench. A more promising

varient, which does not decrease diversity, is suggested by sexual species where genes are swapped during mating and competition occurs on a smaller scale between genes instead of individuals. This might translate to an optimisation strategy (method D) similar to method C but instead of discarding all of the longer tour after τ attempted mutations, the tours are searched for subroutes where some common subset of cities are visited in a different order (see figure 6). The shorter subroute is then copied over to replace the longer one and the quenches continue. One would expect this method to produce shorter tours than A and C; and also to be some improvement on B, which has two independent paths, since some of the best features of both paths are included in D. This is observed. Method D bears some resemblance to the partitioning strategy of Karp [22] for the TSP.
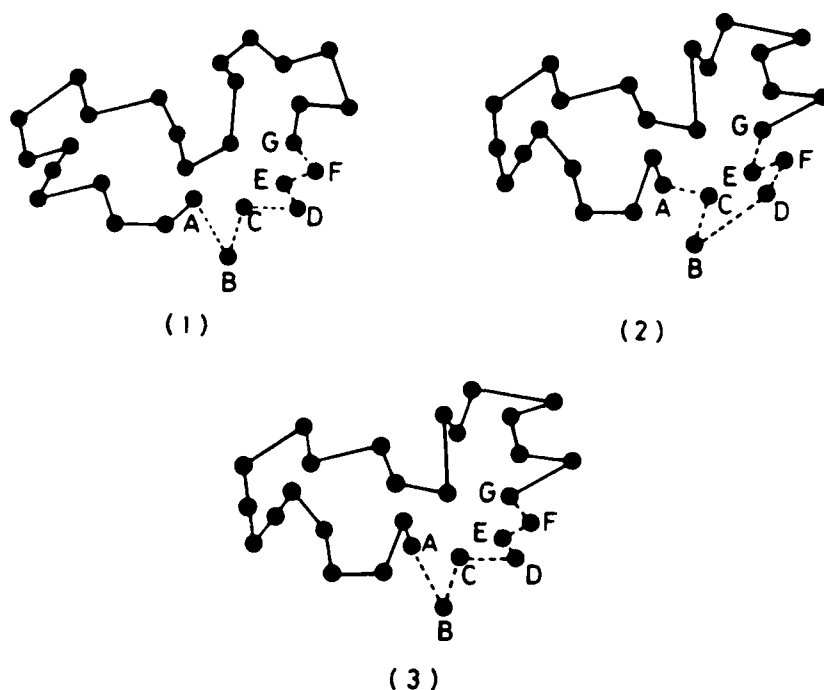

Figure 6: "Gene swapping during mating". (1) and (2) show two independent tours after τ mutations. The subset of cities (A G) occurs in both tours, but they are visited in a different order. Since the subroute ABCDEFG is shorter than ACBDFEG, the dashed segment of tour (2) is replaced with that of tour (1) to give a new tour (3). Tour (2) is discarded and optimisation continues from (1) and (3).

In the methods described above mutations were chosen randomly. The last two methods suggested by Brady utilise the mating strategy of method D but replace the random quench with a systematic search for favourable mutations. In method E, mating is only carried out between two independent tours when each is optimum in the sense that all pair interchanges of cities would give longer tours. Such tours are 2-OPT, and mating provides an escape route from 2-OPT local minima. Finally, method F is very similar to method E but there are twelve independent paths instead of two. Six matings are carried out between randomly chosen partners after all twelve tours are 2-OPT, and the six altered individuals are systematically quenched to new 2-OPT tours before randomly chosen pairs mate again. Method F produces the shortest tours found, and Brady suggests that it is more successful than mating between a single pair of individuals (E) because there is a wider gene pool from which to choose subroutes.

Method F, systematic quenching to local 2-OPT minima followed by mating to escape, is claimed to produce shorter tours than simulated annealing and to require less computer time. However, on the limited numerical data presented so far it is not certain that this is true even for the 64 city TSP. The calculations consist of 100 different runs for each method (with the exception of the OSA algorithm for which there are only 20) for a single configuration of 64 cities. To reach any firm conclusions would require averaging not only over a larger number of runs for a given instance of the 64 city TSP, but over many different instances as well. In spin glass terms, it is vital to average over bond sets $\{J_{ij}\}$ in order to get quantitative information. Furthermore, even if one of these methods does perform better than OSA for 64 cities, it is not certain to do so for larger problems, as noted by Bonomi and Lutton for other heuristic algorithms. Indeed, since the most successful of Brady's methods require finding 2-OPT local minima, and the number of 2-OPT tours grows as $O(N^2)$, this may well be the case.

Notwithstanding these comments, Brady has produced an interesting new approach to optimisation which may yield useful algorithms, particularly for some classes of parallel computers.

4. OPTIMISATION BY "NEURAL NETWORKS"

It has often been suggested that optimisation problems form an important part of biological perception tasks such as the early pattern recognition stages of vision and speech perception. In view of the massive amounts of data to be processed at any instant, and the rapidity with which solutions to such highly complicated problems are found, it is clear that biological information processing systems are fast, powerful computers for pattern recognition tasks. Can one make highly simplified models of neural networks which retain this computational power and speed for other optimisation problems?

Despite an enormous amount of neurophysiological research, the mechanisms of biological computations are largely a mystery. Perhaps this is not surprising. One might learn something about the structure of a computer by taking one apart, but little about the algorithms which the computer is able to execute. The biological status is similar: some of the circuits are known, and so are the characteristics of some components, but the computational mechanisms are unclear. However, three key features have emerged.

First, for at least some processes, parts of the neural system are believed to carry out highly parallel computations. Second, the connectivity of neural networks is much higher than in VLSI (Very Large Scale Integration) devices. Each neuron is connected, typically, to $10^2 \rightarrow 10^3$ others, while connections between cells on VLSI chips are usually between nearest neighbours only. The third feature, which has long been known but the significance of which has only recently been exploited, is that the biological system operates in an analog manner. Clusters of neurons adjust their outputs simultaneously and self-consistently in response to the inputs from hundreds or thousands of other neurons. It is thought to be the combination of high connectivity and analog behaviour which provides the speed. As Hopfield and Tank have pointed out in their seminal paper [23], a network of non-linear neural processors (artificial neurons) works in parallel and computes a collective solution on the basis of simultaneous interactions between all the devices. The solution is found within a few timesteps, where time is measured in the characteristic response times of the neurons. For real neurons the characteristic times are in the millisecond range while for electronic "neurons" the time constants could be much faster - perhaps a few nanoseconds. The fact that analog computation is less accurate than digital computation is unimportant for hard optimisation problems for reasons outlined in the introduction.

For simple optimisation problems, i.e. those with convex cost functions, networks of analog processors can find the global optimum. Tank and Hopfield [24] have shown how to design networks to solve several important examples of this type including an A/D (analog/digital) converter; and the use of analog networks for various early vision problems which can be formulated in terms of convex cost functions has been described by Poggio [25]. However, a major step forward was the demonstration by Hopfield and Tank [23] that networks of analog devices with non-linear responses can also find near-optimal solutions for NP-hard optimisation problems such as the TSP.

We shall describe first the general implementation of an analog "neural" network in terms of normal electronic components. The mathematical behaviour of these networks is compatible with gross features of the physiology of real biological neural networks [23]; though they are, of course, much simpler. The basic element (neuron) is an amplifier with a non-linear response function. Indeed, the interesting behaviour of these networks depends upon the non-linearity of the response. Each amplifier j gives an output voltage V_j which depends upon its input voltage u_j :

$$V_j = g_j(u_j) \tag{5}.$$

The form of the response function is shown in figure 7(a), and the basic circuit for a pair of artificial neurons appears in figure 8. Associated with
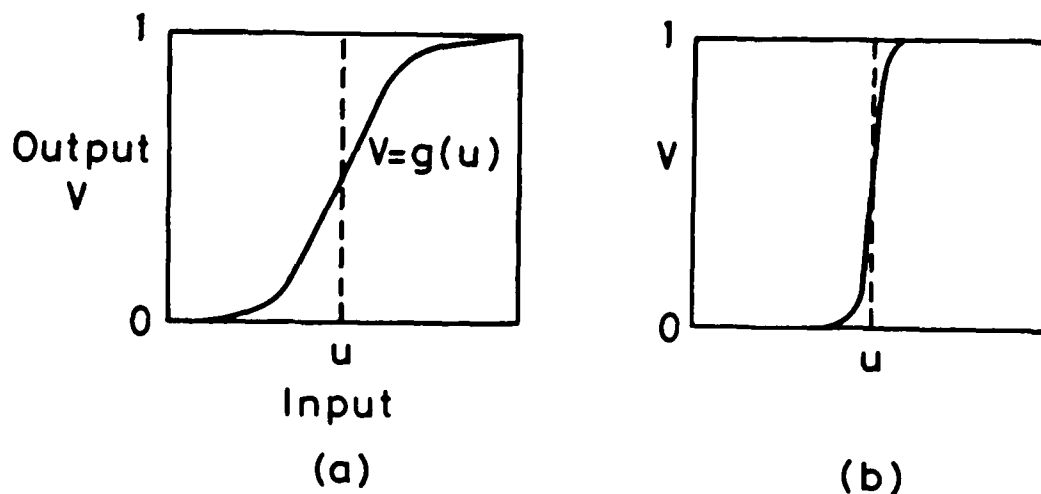
Figure 7: Response function $V = g(u)$ for artificial neurons (amplifiers).
(a) general form – see text for notation; (b) high-gain limit.

▽ Amplifer   ▽ Inverting Amplifer
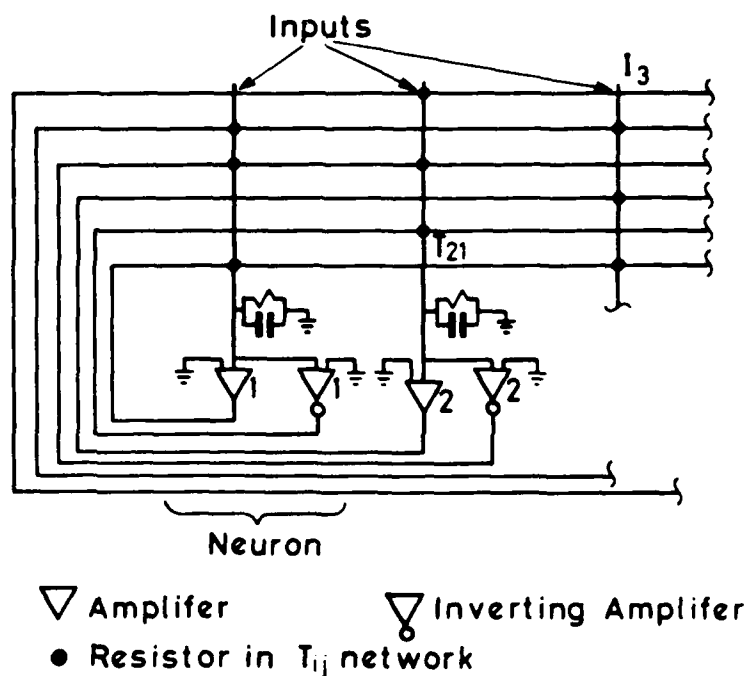● Resistor in T_{ij} network

Figure 8: Basic analog circuit. The output of any neuron can in principle
be connected to the input of any other neuron. Black circles at
intersections represent resistive connections (T_{ij}'s) between
outputs and inputs. Connections between inverted outputs and
inputs represent negative (inhibitory) connections. Reproduced
from [23] with permission.

- 13 -

each amplifier is an input resistor ρ_j leading to a reference ground, and an input capacitor C_j. The synaptic link between neurons i and j is modelled by a conductance T_{ij}. Since real neural networks have both excitatory and inhibitory connections, it is necessary to give each amplifier two outputs: a normal output (V_j in the range $0 \to 1$) and an inverted output (V_j: $0 \to -1$). For excitatory links ($T_{ij} > 0$) the connection is made between the normal output of amplifier j and the input of amplifier i; inhibitory links ($T_{ij} < 0$) are made via the inverted output. Finally, each amplifier has an externally supplied input current I_j which essentially shifts the response curve along the u axis.

The time evolution of such circuits is described by the equations of motion [23]:

$$C_i \, (du_i/dt) = \sum_{j=1}^{N} T_{ij} V_j - (u_i/R_i) + I_i \qquad (6)$$

where

$$\frac{1}{R_i} = \frac{1}{\rho_i} + \sum_{j=1}^{N} \frac{1}{R_{ij}}$$

and $R_{ij} = 1/|T_{ij}|$. For simplicity, all amplifiers can be given the same characteristics, which removes the subscripts from C_i, R_i and g_i. Redefining T_{ij} as T_{ij}/C, and I_i as I_i/C gives a new set of equations of motion:

$$(du_i/dt) = -(u_i/\tau) + \sum_{j=1}^{N} T_{ij} V_j + I_i \qquad (7)$$

where $\tau = RC$, and $V_j = g(u_j)$.

Given some initial set of conditions (the values of $\{u_i\}$ at t=0) integration of equation (7) by any suitable method allows the network to be simulated. If the connections are symmetric, $T_{ji} = T_{ij}$, Hopfield has shown [26] that these networks converge to stable states at which $dV_j/dt = 0$ for all j. Furthermore, in the high gain limit (see figure 7(b)) the stable states are local minima of the function

$$E = -(1/2) \sum_{i=1}^{N} \sum_{j=1}^{N} V_i T_{ij} V_j - \sum_{i=1}^{N} V_i I_i \qquad (8)$$

which is a familiar function in discrete two-state neural network models where $V_i = 0$ or 1 [26]. This is the relation between stable states in the discrete and continuous cases, and shows how a discrete problem can be embedded in a continuous space. The minima of the discrete case occur at the corners of an N dimensional hypercube; the state space for the analog circuit is the interior of the same hypercube; and in the high gain limit the stable states approach the corners of the hypercube.

To use an analog network of the type shown in figure 8 to solve some optimisation problem, one first chooses $\{T_{ij}\}$ and the external input currents $\{I_i\}$ to represent the cost function. Some initial input voltages $\{u_i\}$ are then provided; the network is allowed to evolve to a stable state; and the final state is interpreted as a low-cost configuration.

Hopfield and Tank model the TSP thus. n cities are mapped onto a network of $N=n^2$ neurons. Each city is associated with a vector of n neurons, all except one of which has zero value. The non-zero (=1) value indicates the position of the city in the tour. The whole tour is then described by a matrix. For example, the matrix:

position in tour

		1	2	3	4	5	6
	A	1	0	0	0	0	0
	B	0	1	0	0	0	0
city	C	0	0	0	0	1	0
	D	0	0	0	0	0	1
	E	0	0	1	0	0	0
	F	0	0	0	1	0	0

represents the tour ABEFCD shown in figure 3. (Note that the TSP is inherently discrete if elements of the matrix are identified as output voltages, so the high gain limit is required.) A cost function for this problem must satisfy two conditions:

(1) minima must correspond to permutation matrices like the one shown above;

(2) of these n! valid permutations, the function must have minima which correspond to the few shortest tours.

- 15 -

The first condition can be satisfied by the function

$$E = (A/2) \sum_X \sum_i \sum_{j \neq i} V_{Xi} V_{Xj} + (B/2) \sum_i \sum_X \sum_{Y \neq X} V_{Xi} V_{Yi}$$

$$+ (C/2) \left(\sum_X \sum_i V_{Xi} - n \right)^2 \qquad (9)$$

where A, B and C are positive constants and V_{Xi} is the element of the matrix for city X and position i. This gives a minimum value of E=0 if and only if $\underset{\sim}{V}$ is a permutation matrix because

(i) the first triple sum is zero if and only if each row contains only one unit element (the rest being zero).

(ii) The second triple sum is zero if and only if each column contains a single unit element (ditto).

(iii) The third term is zero if and only if there are n unit elements in the whole matrix.

These three terms strongly favour valid tours: i.e. they embody constraints which apply to any TSP.

To meet the second condition, a fourth data-dependent term is added which describes which particular TSP is to be solved:

$$+ (D/2) \sum_X \sum_{Y \neq X} d_{XY} \sum_i V_{Xi} (V_{Y,i+1} + V_{Y,i-1}) \qquad (10)$$

This term gives E equal to the length of the tour, for valid tours selected by the first condition. Thus if A, B and C are large enough, all low energy states described by eqns (9) and (10) are valid tours, and the value of E is the tour length. Given suitable starting conditions and values for the constants, a network should be able to compute near-optimal tours.

Hopfield and Tank tested this method on a 10 city TSP, and further details can be found in their paper [23]. The evolution of a network to a stable solution is shown in figure 9. For a chosen set of parameters A, B, C, D, and given the initial conditions, the calculation is deterministic. However, slightly different initial conditions lead to different solutions. In 20 simulations starting from different initial states, Hopfield and Tank
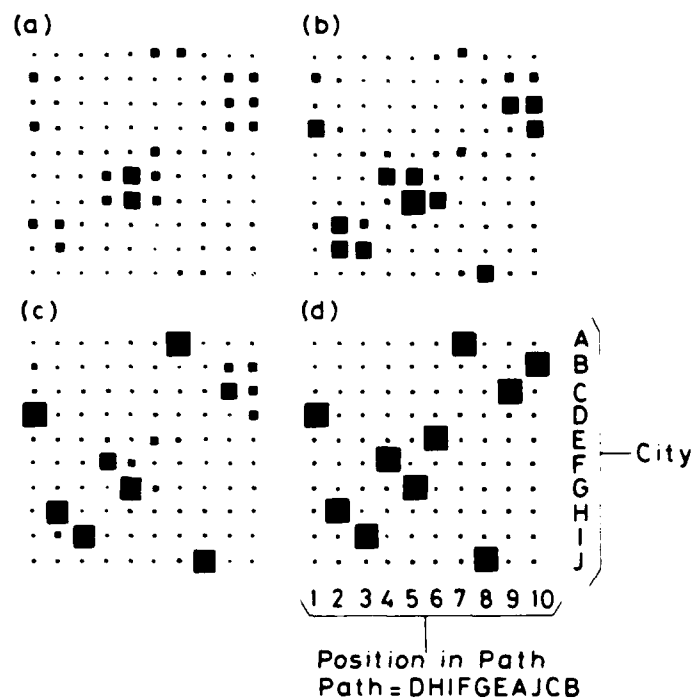
Figure 9: Hopfield and Tank's TSP results [23]. This shows the convergence
of the 10-city analog circuit to a tour. The linear dimension
of each square is proportional to the value of V_{Xi}. (a), (b)
and (c) show intermediate times; (d) is the final state. The
indices in (d) show how the final state is decoded into a tour.
Reproduced by permission from [23].

found that 16 converged to valid tours, and about half of the trials produced
one of the two shortest tours found by exhaustive search. This is impressive
performance: for 10 cities there are $(n-1)!/2 = 181,440$ different tours.
Furthermore, convergence to a stable solution is rapid in term of the time
constant τ.

Clearly, future applications of the method will depend on building
circuits in hardware rather than solving the corresponding differential
equations on a serial computer. A key question is therefore how the behaviour
of the network scales with size. One potential problem is that in simulations
of a 900 neuron network for the 30 city TSP, Hopfield and Tank found that the
choice of parameters was more critical than in the 100 neuron network. It
will require much simulation work, and building real circuits, before the
scaling behaviour of these systems for model problems such as the TSP is
understood in detail. There is also much work to be done in order to discover
the range of computations for which such networks are efficient. However, it
seems plausible that networks of simple analog devices have the potential to
carry out powerful optimisations more quickly than any other method known.

5. SUMMARY

In this paper I have described several novel approaches to combinatorial optimisation which have been prompted by models of natural phenomena. All are in their infancy; and each has much to offer. Optimisation by Simulated Annealing has already been established as a powerful tool for many practical problems. There are basic unsolved questions, about the choice of near-optimal annealing schedules for practical applications, which should exercise mathematicians!. Apart from such immediate matters, the close links with spin glass physics promises new insights into what makes some optimisation problems hard; and possibly a new classification of computational complexity. Brady's approach, or other strategies gleaned from evolution, may or may not prove to be useful optimisation tools: further calculations are needed. However, such lateral thinking may well stimulate further developments. The most exciting prospect of all is the possibility that we may be able to build novel sorts of computers, based on networks of relatively simple non-linear analog devices, which will be capable of quite different computations from those performed by digital computers. Will the next few years see the birth of the Hopfield computer?

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Toulouse, G., Comm. Phys., 2 (1977) 115-119.

[2] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A.,
    and Teller,E., J. Chem. Phys., 21 (1953) 1087-1092.

[3] Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P., Science, 220 (1983)
    671-680.

[4] Černý, V., J. Optimisation Theory Appl. 45 (1985) 41-51.

[5] Pincus, M., Op. Res., 18 (1970) 1225-1228.

[6] Ackley, D.H., Hinton, G.E., and Sejnowski, T.J., Cognitive Science,
    9 (1985) 147-168.

[7] Kirkpatrick, S., J. Stat. Phys., 34 (1984) 975-986.

[8] Geman, S. and Geman, D., IEEE PAMI 6 (1984) 721-741.

[9] Gidas, B., J. Stat. Phys., 39 (1985) 73-131.

[10] Kirkpatrick, S. and Toulouse, G., J. Physique 46 (1985) 1277-1292.

[11] Bonomi, E. and Lutton, J., SIAM Rev., 26 (1984) 551-567.

[12] Lin, S., Bell Syst. Tech. J., 44 (1965) 2245-2269.

[13] Bachas, C.P., J. Phys. A. 17 (1984) L709-L712.

[14] Barahona, F., Maynard, R., Rammal, R. and Uhry, J.P., J. Phys. A. 15
    (1982) 673.

[15] Orland, H., J. Physique Lett., 46 (1985) L763-L770.

[16] Edwards, S.F. and Anderson, P.W., J. Phys. F. Metal Phys.,
    5 (1975) 965-974.

[17] Mezard, M. and Parisi, G., J. Physique Lett. 46, (1985) L771-L778.

[18] Mezard, M. and Parisi, G., "A replica analysis of the travelling salesman
    problem", preprint, 1986.

[19] Vannimenus, J. and Mezard, M., J. Physique Lett., 45 (1984)
    L1145-L1153.

[20] Angles d'Auriac, J.C., Preissmann, M. and Rammal, R., J. Physique Lett.
    46 (1985) L173-L180.

[21] Brady, R.M., Nature, 317 (1985) 804-806.

[22] Karp, R.M., Math. Oper. Res., 2 (1977) 209-223.

[23] Hopfield, J.J. and Tank, D.W., Biol. Cyb. 52 (1985) 141-152.

[24] Tank, D.W. and Hopfield, J.J., IEEE Circuits and Systems, 1986, in
    press.

[25] Poggio, T., Comp. Vis. Graph. Image Proc., 31 (1985) 139-155;
    Poggio, T., Torre, V. and Koch, C., Nature 317 (1985) 314-319.

[26] Hopfield, J.J., Proc. Nat. Acad. Sci. USA 81 (1984) 3088-3092.

END

DTIC

8-86